

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-199332

(43)Date of publication of application : 12.07.2002

(51)Int.CI.

H04N 5/91  
G06T 7/00  
G06T 7/20  
G11B 20/10

(21)Application number : 2000-396820

(71)Applicant : KDDI CORP

(22)Date of filing : 27.12.2000

(72)Inventor : SUGANO MASARU

NAKAJIMA YASUYUKI

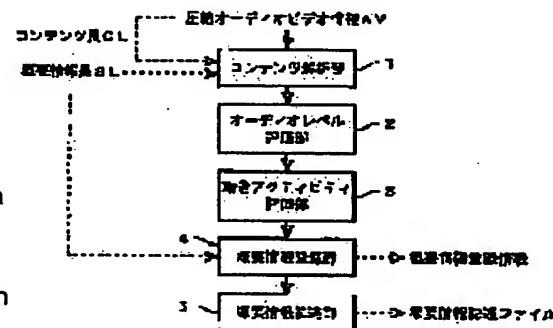
YANAGIHARA HIROMASA

## (54) AUDIO SUMMARY INFORMATION, AND EXTRACTION DEVICE THEREFOR, REPRODUCTION DEVICE AND RECORDING MEDIUM

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To provide audio summary information and the extraction device therefor, the reproduction device and the recording medium, which extracts summary information having a length designated by audio or audio video information uniformly and speedily from the whole contents.

**SOLUTION:** A contents analytic part 1 analyzes the hourly structure of video information of inputted audio video contents. An audio level evaluation part 2 evaluates the audio level of audio information attached to the video information the hourly structure of which is analyzed. A movement activity evaluation part 3 evaluates the movement activity of a shot which is evaluated to be the candidate of the summary information by the part 2. A summary information registering part 4 registers audio video summary information evaluated by the movement activity evaluation part.



### LEGAL STATUS

[Date of request for examination] 10.02.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-199332

(P2002-199332A)

(43)公開日 平成14年7月12日 (2002.7.12)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テ-コ-ト*(参考)
H 0 4 N 5/91		G 0 6 T 7/00	3 0 0 F 5 C 0 5 3
G 0 6 T 7/00	3 0 0	7/20	B 5 D 0 4 4
7/20		G 1 1 B 20/10	G 5 L 0 9 6
G 1 1 B 20/10		H 0 4 N 5/91	N
			C

審査請求 未請求 請求項の数28 OL (全12頁)

(21)出願番号 特願2000-396820(P2000-396820)

(71)出願人 000208891

ケイディーディーアイ株式会社

東京都新宿区西新宿二丁目3番2号

(22)出願日 平成12年12月27日 (2000.12.27)

(72)発明者 皆野 勝

埼玉県上福岡市大原2-1-15 株式会社

ケイディディ研究所内

(72)発明者 中島 康之

埼玉県上福岡市大原2-1-15 株式会社

ケイディディ研究所内

(74)代理人 100084870

弁理士 田中 香樹 (外1名)

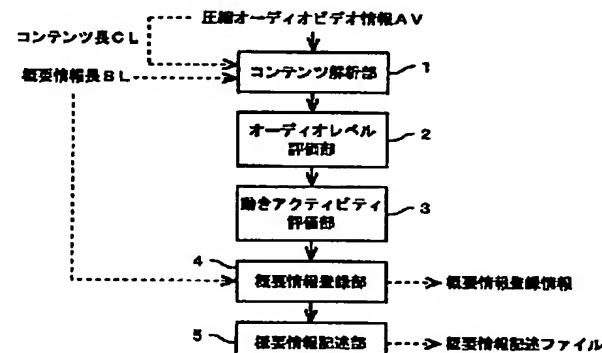
最終頁に続く

(54)【発明の名称】 オーディオ概要情報、オーディオビデオ概要情報の抽出装置、再生装置、および記録媒体

(57)【要約】

【課題】オーディオまたはオーディオビデオ情報の指定した長さを持つ概要情報を、コンテンツ全体から均一に、かつ高速に抽出することを可能とした、オーディオ概要情報、オーディオビデオ概要情報の抽出装置、再生装置、および記録媒体を提供することにある。

【解決手段】入力されたオーディオビデオコンテンツのビデオ情報の時間的構造を解析するコンテンツ解析部1と、該時間的構造を解析された該ビデオ情報に付随するオーディオ情報のオーディオレベルを評価するオーディオレベル評価部2と、該オーディオレベル評価部2で概要情報の候補とされたショットの動きアクティビティを評価する動きアクティビティ評価部3と、該動きアクティビティ評価部により評価されたオーディオビデオ概要情報を登録する概要情報登録部4とからなる。



1

## 【特許請求の範囲】

【請求項1】 入力されたオーディオビデオコンテンツのビデオ情報の時間的構造を解析するコンテンツ解析手段と、  
時間的構造を解析された該ビデオ情報に付随するオーディオ情報のオーディオレベルを評価するオーディオレベル評価手段と、  
該オーディオレベル評価手段により評価されたオーディオビデオ概要情報を登録する概要情報登録手段と、具備したことと特徴とするオーディオビデオ概要情報抽出装置。

【請求項2】 請求項1に記載のオーディオビデオ概要情報抽出装置において、

前記コンテンツ解析手段は、該ビデオ情報をショットに分割する手段と、入力コンテンツをある基準に従って等間隔の区間に分割する手段と、該等間隔の区間に含まれるショットの長さを評価する手段および反復ショットを判定する手段の少なくとも一方とを具備し、  
該ショットの長さが予め定められた値より小さい時または反復ショットと判定された時に、該ショットを概要情報から除外するようにしたことを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項3】 請求項2に記載のオーディオビデオ概要情報の抽出装置において、

前記反復ショットを判定する手段は、前記分割されたショットの代表フレームを決定する手段と、該代表フレームの特徴値を抽出する手段と、該特徴値と他の代表フレームの特徴値とを比べて類似画像検索を行う手段とを具備し、  
該類似画像検索により類似画像と判断された時に、該ショットを反復ショットとして除外することを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項4】 請求項2に記載のオーディオビデオ概要情報の抽出装置において、

前記反復ショットを判定する手段は、前記分割されたショットの特徴値を抽出する手段と、該特徴値と他のショット特徴値とを比べて類似ショット検索を行う手段とを具備し、

該類似ショット検索により類似画像と判断された時に、該ショットを反復ショットと判定することを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項5】 請求項2に記載のオーディオビデオ概要情報の抽出装置において、

前記オーディオレベル評価手段は、無音を判定する手段および低レベル音を判定する手段の少なくとも1つを具備し、  
無音または低レベル音と評価された時には、前記ショットを概要情報から除外することを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項6】 請求項5に記載のオーディオビデオ概要

2

## 情報の抽出装置において、

高レベル音を判定する手段をさらに具備し、該高レベル音を判定する手段は、圧縮されたオーディオ情報からサブバンドデータを抽出する手段と、該サブバンドデータからオーディオのレベルを推定する手段と、ある閾値を超える該オーディオレベルからオーディオレベルの時間密度を計算する手段とを具備し、  
該オーディオレベル時間密度を用いて概要情報の候補を判定することを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項7】 請求項5に記載のオーディオビデオ概要情報の抽出装置において、

高レベル音を判定する手段をさらに具備し、該高レベル音を判定する手段は、圧縮されたオーディオ情報からサブバンドデータを抽出する手段と、該サブバンドデータからバンドにより重み付けしたサブバンドエネルギーの総和を計算する手段と、単位時間におけるサブバンドエネルギー総和を計算する手段とを具備し、

前記重み付けしたサブバンドエネルギーの総和および前記単位時間におけるサブバンドエネルギー総和の少なくとも1つを用いて概要情報の候補を判定することを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項8】 請求項1ないし7のいずれかに記載のオーディオビデオ概要情報の抽出装置において、

画像内の動きを評価する動きアクティビティ評価手段をさらに具備し、

前記概要情報登録手段は、前記オーディオレベル評価手段と前記動きアクティビティ評価手段により評価されたオーディオビデオ概要情報を登録することを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項9】 請求項8に記載のオーディオビデオ概要情報の抽出装置において、

前記動きアクティビティ評価手段は、前記ショットに属するフレームの動きベクトルデータを抽出する手段と、該抽出された動きベクトルデータから動きアクティビティを計算する手段と、単位時間における動きアクティビティを計算する手段とを具備し、該単位時間における動きアクティビティを用いて概要情報の候補を判定することを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項10】 請求項1ないし9のいずれかに記載のオーディオビデオ概要情報の抽出装置において、  
前記概要情報の判定基準となる閾値を再帰的に制御することにより、外部から指定された概要情報長になるまで判定処理を繰り返すことを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項11】 請求項1ないし10のいずれかに記載のオーディオビデオ概要情報の抽出装置において、  
抽出された概要情報の時間情報として、該概要情報の開始時間と終了時間、または開始時間と継続時間を少なくとも記述する手段を具備したことを特徴とするオーディ

オーディオ概要情報の抽出装置。

【請求項12】 請求項1ないし10のいずれかに記載のオーディオビデオ概要情報の抽出装置において、抽出された概要情報を、オーディオビデオとして結合するか、またはオーディオとビデオ個別に結合する手段と、該結合した概要情報を別ファイルとして出力し保存する手段とを具備したことを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項13】 請求項1ないし10のいずれかに記載のオーディオビデオ概要情報の抽出装置において、抽出された概要情報のビデオ要素としてショットの先頭フレームまたは特徴フレームを代表フレームとして抽出する手段と、該概要情報のオーディオ要素としてショットに付随するオーディオ情報を抽出する手段を具備したことを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項14】 請求項13に記載のオーディオビデオ概要情報の抽出装置において、該抽出されたビデオ要素の時間情報としてコンテンツ内の時間位置を記述する手段と、該抽出されたオーディオ要素の時間情報としてコンテンツ内の開始時間と終了時間、または開始時間と継続時間を少なくとも記述する手段を具備したことを特徴とするオーディオビデオ概要情報の抽出装置。

【請求項15】 入力された圧縮オーディオコンテンツからサブバンドデータを抽出する手段と、該サブバンドデータから、バンドにより重み付けしたサブバンドエネルギーの総和を計算する手段と、単位時間におけるサブバンドエネルギー総和を評価する手段と、オーディオ概要情報を抽出する手段とを具備したことを特徴とするオーディオ概要情報の抽出装置。

【請求項16】 入力された圧縮オーディオコンテンツからサブバンドデータを抽出する手段と、該サブバンドデータから、バンドにより重み付けしたサブバンドエネルギーの総和を計算する手段と、該サブバンドエネルギー総和のある単位時間における総和を計算する手段と、概要情報の開始時間を判定する手段と、

概要情報の終了時間を判定する手段と、該開始時間及び終了時間の区間のオーディオ情報を概要情報として抽出する手段を具備したことを特徴とするオーディオ概要情報の抽出装置。

【請求項17】 入力されたオーディオコンテンツをサブバンドデータに変換する手段と、該サブバンドデータから、バンドにより重み付けしたサブバンドエネルギーの総和を計算する手段と、時間単位でのサブバンドエネルギー総和を評価する手段と、

オーディオ概要情報を抽出する手段を具備したことを特

徴とするオーディオ概要情報の抽出装置。

【請求項18】 入力されたオーディオコンテンツをサブバンドデータに変換する手段と、該サブバンドデータから、バンドにより重み付けしたサブバンドエネルギーの総和を計算する手段と、該サブバンドエネルギー総和のある単位時間における総和を計算する手段と、概要情報の開始時間を判定する手段と、概要情報の終了時間を判定する手段と、

10 該開始時間及び終了時間の区間のオーディオ情報を概要情報として抽出する手段を具備したことを特徴とするオーディオ概要情報の抽出装置。

【請求項19】 請求項15ないし18のいずれかに記載のオーディオ概要情報の抽出装置において、抽出された概要情報の時間情報として、該概要情報の開始時間と終了時間、または開始時間と継続時間を少なくとも記述する手段を具備したことを特徴とするオーディオ概要情報の抽出装置。

【請求項20】 請求項15ないし18のいずれかに記載のオーディオ概要情報の抽出装置において、抽出された概要情報を個別のファイルとして出力し保存する手段を具備したことを特徴とするオーディオ概要情報の抽出装置。

【請求項21】 オーディオビデオ概要情報の再生装置において、請求項13に記載のオーディオビデオ概要情報の抽出装置により抽出されたビデオ要素とオーディオ要素を同期して再生する手段を具備したことを特徴とするオーディオビデオ概要情報の再生装置。

【請求項22】 オーディオビデオ概要情報の再生装置において、

請求項13に記載のオーディオビデオ概要情報の抽出装置により抽出されたオーディオ要素を時間的に間引いてオーディオの再生速度を変換する手段と、該変換されたオーディオ要素とビデオ要素を同期して再生する手段を具備したことを特徴とするオーディオビデオ概要情報の再生装置。

【請求項23】 オーディオビデオ概要情報の再生装置において、

40 抽出された概要情報としてのショットのビデオ情報を時間的に間引いてビデオの再生速度を変換する手段と、該抽出された概要情報としてのショットのオーディオ情報をビデオ情報と同一の割合で時間的に間引いてオーディオの再生速度を変換する手段と、該再生速度を変換されたビデオ情報とオーディオ情報を同期して再生する手段を具備したことを特徴とするオーディオビデオ概要情報の再生装置。

【請求項24】 請求項22または23に記載のオーディオビデオ概要情報の再生装置において、前記オーディオの再生速度変換は、オーディオのフレー

5

ムを周期的にスキップすることにより達成されることを特徴とするオーディオビデオ概要情報の再生装置。

【請求項25】 請求項22または23に記載のオーディオビデオ概要情報の再生装置において、

前記オーディオの再生速度変換は、オーディオのフレームの繰り返し再生と周期的なスキップを組み合わせることにより達成されることを特徴とするオーディオビデオ概要情報の再生装置。

【請求項26】 入力されたオーディオビデオコンテンツのビデオ情報の時間的構造を解析する機能と、

時間的構造を解析された該ビデオ情報に付随するオーディオ情報のオーディオレベルを評価する機能と、

前記オーディオレベルと動きアクティビティを評価されたオーディオビデオ概要情報を登録する概要情報登録機能とを含む、

コンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項27】 請求項26に記載の記録媒体において、

画像内の動きアクティビティを評価する機能をさらに具備し、前記概要情報登録機能は、前記オーディオレベルと動きアクティビティを評価されたオーディオビデオ概要情報を登録することを特徴とする記録媒体。

【請求項28】 入力された圧縮オーディオコンテンツからサブバンドデータを抽出するサブバンドデータ抽出機能および入力されたオーディオコンテンツをサブバンドデータに変換するサブバンド解析機能の少なくとも一方と、

高レベル音を評価する機能と、

前記オーディオコンテンツの概要情報を登録する機能とを含む、

コンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、非圧縮または圧縮されたオーディオ情報、または圧縮されたオーディオビデオ情報から、それらの内容を効率的に把握するための概要情報（サマリ）を抽出するオーディオ情報、オーディオビデオ概要情報の抽出装置および記録媒体に関する。また、非圧縮または圧縮されたオーディオ情報、または圧縮されたオーディオビデオ情報から圧縮データ領域で高速に概要情報を抽出することにより、オーディオ情報またはオーディオビデオ情報の高速かつ効率的な閲覧を提供することが可能な、オーディオビデオ概要情報の再生装置に関する。

【0002】

【従来の技術】ビデオの自動要約作成については、例えば田中、脇本、神田による「シーン検出による動画情報の自動要約・閲覧技術の開発」、電子情報通信学会技術

10

報告、IE99-20（1999）で研究発表（第1の従来技術）されており、該研究発表では、ビデオのシーン変化点を検出した後、階層構造化を行い、各シーンに優先度を付与することによってビデオの要約を自動的に作成している。特に、話題の転換点直後のシーンには高い優先度が付与される。階層構造においては、上位階層ほど優先度が高くなるように設定される。

【0003】また、J.Saarela、B.Merialdoによる「Using content models to build audio-video summaries」、SPIE Conference on Storage and Retrieval for Imageand Video Databases VII (1999)においては（第2の従来技術）、汎用的なビデオの概要情報（サマリ）の作成を、制約つきの最適化問題として捉えている。制約としては、最小のショット長、オーディオとビデオの同期、ビデオの連續性、及びオーディオとビデオの冗長性などである。そして、手動によってビデオの内容モデル（記号的な記述）を構築し、サマリの作成を行っている。

【0004】また、R.Lienhartらによる「Video Abstracting」、Communications of ACM, Vol.40, No.12 (1997) では（第3の従来技術）、映画の予告編に特化した概要情報の作成を目的としている。主な作成手順としては、ビデオのショットへの分割、特別なイベントを含むクリップの解析、クリップの選択、およびクリップの集約である。特別なイベントとしては、俳優の顔の認識・会話の識別、タイトルからの文字情報の抽出、及び銃撃や爆発などのイベントを抽出している。

【0005】

【発明が解決しようとする課題】これらの従来技術によるオーディオビデオの概要情報抽出は、次のような問題を有している。まず、前記第1の従来技術においては、ビデオの階層構造化により効率的な要約作成を提供するが、階層構造化は手動で行う必要があり、長尺のオーディオビデオ情報に関しては概要情報の抽出に必要な前処理としての階層構造化に、多くの時間を要する可能性がある。また、優先度の付与はシーンが属する階層に依存して行われるため、実質的には人手を要することが多くなる。また、ビデオ単体が対象となっているためオーディオビデオへの拡張は可能であるが、オーディオとしての特性を利用していないため、場合によっては重要な内容を含むオーディオ情報を利用していないため、適切な概要情報が得られないことも考えられる。

【0006】また、前記第2の従来技術においては、ビデオの内容モデルの構築を手動で行わなければならないほか、効果的にオーディオの特性を利用する方式を探っていない。同時に、ビデオのセグメントの分類においては、圧縮データ上では実現が困難な、高度な認識技術などが必要となるため、概要情報抽出に要する処理が大きくなることが予想される。前記第3の従来技術についても、これらの高度な（コンテンツの意味内容にまで立ち

50

7

入った) 处理が必要となっている。

【0007】このように、従来技術ではオーディオビデオ情報の入力から概要情報の出力までの過程において、手動による処理が介在することが多く、また、オーディオ情報を効果的に利用しないため、オーディオとしての特性による概要情報の作成が考慮されていない。また、圧縮または非圧縮のオーディオ単体からの自動概要情報抽出については、これまで有効な方式は検討されていない。

【0008】また、概要情報の構造的な側面から見て  
も、オーディオビデオ情報全体から均一に概要情報が抽出  
される保証はなく、さらに前記第1と第2の従来技術  
では、外部から指定された概要情報長に近づけるよう  
な制御を行う方式も十分には採用されていない。

【0009】本発明は、前記した従来技術に鑑みてなされたものであり、その目的は、非圧縮または圧縮されたオーディオ情報、または圧縮されたオーディオビデオ情報から、それらの内容を効率よく把握するための概要情報（サマリ）を抽出する装置において、圧縮データ領域でオーディオ及びビデオの時空間的な特性を解析し、必要に応じてこれらを統合的に評価し、オーディオまたはオーディオビデオ情報の指定した長さを持つ概要情報を、コンテンツ全体から均一に、かつ高速に抽出することを可能とした、オーディオ概要情報、オーディオビデオ概要情報の抽出装置、再生装置および記録媒体を提供することにある。

{0010}

【課題を解決するための手段】前記した目的を達成するために、本発明は、オーディオビデオ概要情報の抽出装置において、入力されたオーディオビデオコンテンツのビデオ情報の時間的構造を解析するコンテンツ解析手段と、時間的構造を解析された該ビデオ情報に付随するオーディオ情報のオーディオレベルを評価するオーディオレベル評価手段と、該オーディオレベル評価手段により評価されたオーディオビデオ概要情報を登録する概要情報登録手段とを具備した点に第1の特徴がある。

【0011】また、本発明は、オーディオ概要情報の抽出装置において、入力された圧縮オーディオコンテンツからサブバンドデータを抽出する手段と、該サブバンドデータから、バンドにより重み付けしたサブバンドエネルギーの総和を計算する手段と、単位時間におけるサブバンドエネルギー総和を評価する手段と、オーディオ概要情報を抽出する手段とを具備した点に第2の特徴がある。

【0012】また、本発明は、入力されたオーディオビデオコンテンツのビデオ情報の時間的構造を解析する機能と、時間的構造を解析された該ビデオ情報に付随するオーディオ情報のオーディオレベルを評価する機能と、画像内の動きアクティビティを評価する機能と、前記オーディオレベルと動きアクティビティを評価されたオー

8

ディオビデオ概要情報を登録する概要情報登録機能とを含む、コンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体を提供するようにした点に第3の特徴がある。

【0013】さらに、本発明は、オーディオビデオ概要情報の再生装置において、オーディオビデオ概要情報の抽出装置により抽出されたビデオ要素とオーディオ要素を同期して再生する手段を具備した点に第4の特徴がある。

10 [0014]

【発明の実施の形態】以下に、図面を参照して、本発明を詳細に説明する。図1は、本発明の第1の実施形態であるオーディオビデオ概要情報の抽出装置の構成を示すブロック図である。

【0015】まず、圧縮されたオーディオビデオ情報AVは、該オーディオビデオ情報AVの時間的構造を解析するコンテンツ解析部1に入力される。該コンテンツ解析部1は、図2に示されているような構成を有しており、該圧縮されたオーディオビデオ情報AVは、まずショット分割部11に入力する。該ショット分割部11は、ビデオ情報をショット単位Sに分割し、次いでコンテンツ全体に対するショット数をカウントし、全ショット数をNSとして保持する。次に、分割数決定部12は、ショット分割部11から入力される全ショット数NSと、オーディオビデオ情報AVから得られるコンテンツ長CLと、外部から指定される概要情報長SLを用いて、下記の(1)式により分割数NPを決定する。

$$[0016] \text{NP} = \text{NS} \times \text{SL} / \text{CL} \quad \dots (1)$$

ただし、SL>CL／NSであるとする。例えば、コンテンツ長CLが1時間のものを5分の概要情報長SLに纏めようすると、前記全ショット数NSが240個の場合、分割数NPは20となる。そして、コンテンツ分割13においてコンテンツをNP等分する。等分割された区間SAVと、各区間に属するショットSとをマッピングする。

【0017】以下では、分割区間入力14で入力された区間SAV毎に処理を行う。まず、ショット入力15で最初のショットSnが入力される。ショット長評価部16では、入力されたショットSnのショット長SHLnが予め指定した長さT以下の場合に、概要情報SUMの候補から除外し、次のショットSn+1の入力処理へ移行する。一方、ショットSnのショット長SHLnがTよりも長い場合には、該当するショットSnを概要情報SUMの候補とし、代表フレーム抽出部17に送る。

【0018】代表フレーム抽出部17では、ショットSnの代表フレームRFSnとして、ショットの先頭フレームSF Snまたは特徴フレームKFSnを抽出する。特徴フレームKF Snの抽出には、例えば特願2000-065259に記載された方法などを用いることができる。さらに、代表フレーム特徴値抽出部18において、代表フレーム抽出部17で抽

50

出されたフレームRFSnの特徴値CHFSnを抽出し、ショット特徴値抽出部19において、ショットSnの特徴値CHSnを抽出する。

【0019】ショットSnに対する代表フレームRFSnとしての先頭フレームSFSnまたは特徴フレームKFSnの特徴値CHFSnと、ショットSnの特徴値CHSnのいずれか一方又は両方は、代表ショット評価部1Aに送られる。先頭フレームSFSnまたは特徴フレームKFSnの特徴値CHFSnとしては、例えばMPEG-7(Moving Picture Experts Group phase 7)で規定されている記述子などを用いることができる。

【0020】代表ショット評価部1Aでは、対象となる分割区間SAVにおいて既に代表ショットRSとして登録されている全てのショットに関する代表フレームRFの特徴値CHFと、代表フレーム特徴値抽出部18から送られたショットSnの代表フレームRFSの特徴値CHFSnとの間で類似度の判定を行う。ここで、特徴値CHFと入力されたショットSnの特徴値CHFSnとの類似度が大きいと判定された場合には反復ショットであると見なされ、該当するショットSnを概要情報SUMの候補から除外し、次のショットSn+1の入力処理へ移行する。一方、特徴値CHFとショットSnの特徴値CHFSnとの類似度が小さいと判定された場合には、独立ショットであると見なされ、該当するショットSnを概要情報SUMの候補とし、代表ショット登録1Bでショットの登録を行ったあと、オーディオレベル評価部2に送る。なお、代表ショットとして登録されたショットSnに関する特徴値は別途保持され、それ以降の代表ショットの評価に用いられる。

【0021】ショット類似度の判定においては、同様に既に代表ショットRSとして登録されているショットに関する代表ショットの特徴値CHSと、ショット特徴値抽出部19から送られたショットSnの特徴値CHSnを代用するか、或いは併用してもよい。

【0022】オーディオレベル評価部2では、コンテンツ解析部1から入力されたショットSnに関して、図3のサブバンドデータ抽出部21でショットSnのオーディオ部分のサブバンドデータSDSnを抽出する。そして、無音解析部22で無音としての特徴値を計算したあと、無音判定部23でショットSnの全ての区間、またはショットSnの区間のY%以上が無音であると判定された場合に、該当するショットSnを概要情報SUMの候補から除外し、次のショットSn+1の入力処理へ移行する。一方、無音判定部23での判定において、ショットSnのオーディオ部分がショットSnの区間の(100-Y)%以上が有音である場合に、ショットSnは低レベル音解析部24へ送られる。無音解析部22での無音解析方法及び無音判定部23での無音判定方法としては、例えば特願平10-235543号に記載された方法などを用いることができる。

【0023】低レベル音解析部24では、同様にサブバンドデータ抽出部21で抽出されたサブバンドデータSD

Snから該当するオーディオのレベルLSnを推定し、指定された十分に低いレベルTHLL以下のオーディオがショットSnの全ての区間、またはショットSnの区間のZ%以上を占める場合に、該当するショットSnを概要情報SUMの候補から除外し、次のショットSn+1の入力処理へ移行する。一方、低レベル音解析部24でショットSnのオーディオ部分が、ショットSnの区間の(100-Z)%以上がオーディオレベルTHLLを超えるオーディオである場合に、該当するショットSnを概要情報SUMの候補とし、高レベル音解析部26に送る。

【0024】高レベル音解析部26では、サブバンドデータ抽出部21から得られたショットSnに関するサブバンドデータSDSnに基づいて、図4の単位時間密度計算部261でショットSnにおけるあるレベルTHHL以上のオーディオの単位時間密度Dsnを計算する。このとき、オーディオ情報がMPEGオーディオで符号化されている場合、1秒当たりの単位時間密度Dsnは、例えば以下のように求めることができる。

【0025】

20  $D_{sn} = (NAF_{THHL}/NAF) \times AAL_{THHL} \dots (2)$   
ここで、NAF<sub>THHL</sub>はレベルTHHL以上を持つ1秒間当たりのフレーム数、NAFは1秒当たりのオーディオフレーム数、AAL<sub>THHL</sub>はNAF<sub>THHL</sub>に対する平均レベルである。

【0026】また、サブバンドエネルギー総和計算部262においてサブバンドにより重み付けされたサブバンドエネルギーの総和SBEを計算し、それに基づいて単位時間サブバンドエネルギー計算部263で単位時間でのサブバンドエネルギー総和SBSnを計算する。1秒当たりの単位時間サブバンドエネルギーSBSnは、例えば以下の(3)式のように求めることができる。

$$SBE = \sum_{k=0}^{31} \alpha_k \times sb_k \quad (3)$$

$$SBSn = \sum_{k=0}^{NAF} SBE$$

ここで、 $\alpha_k$ はサブバンドkに対する重み付け、 $sb_k$ はあるフレームにおけるサブバンドkのエネルギーである。

【0027】次に、単位時間密度判定部264で該当するショットSnがある閾値THDを超える単位時間密度Dsnを持つ場合に、単位時間サブバンドエネルギー判定部265へ移行する。単位時間サブバンドエネルギー判定部265では、閾値THSBを超える単位時間サブバンドエネルギーSBSnが存在する場合、該当するオーディオを含むショットSnを概要情報の候補として判定し、高レベル音解析ルーチンを抜け出し、動きアクティビティ評価部3へ移行する。

【0028】これに対し、オーディオレベルTHHL以上のオーディオの単位時間密度Dsnが閾値THD未満の場合、或いは単位時間サブバンドエネルギーSBSnが閾値THSB未満の場合、該当するオーディオを含むショットSnを概要情

11

報の候補から除外し、次のショット $S_{n+1}$ の入力処理へ移行する。

【0029】なお、図3の構成において、無音解析部22、低レベル音解析部24、および高レベル音解析部26は、必ずしも全部は必要でなく、少なくとも1つを備えておれば良い。

【0030】動きアクティビティ評価部3では、オーディオレベル評価部2から入力されたショット $S_n$ に関して、図5の動きベクトル抽出部31でショット $S_n$ に属する全てのフレームの動きベクトル情報MVを抽出する。そして、動きアクティビティ計算部32において、動きベクトル情報MVを用いてショット $S_n$ 全体としての動きアクティビティ $MAS_n$ を計算し、それを用いて単位時間動きアクティビティ計算部33において、ある単位時間（例えば1秒、1フレームなど）における動きアクティビティMAを計算する。動きアクティビティ計算部32及び単位時間動きアクティビティ33における処理は、例えばMPEG符号化されたビデオに対して以下のように表せる。ここでは1秒当たりの動きアクティビティと仮定する。

$$\begin{aligned} ASMV &= \sum |MV| \quad |MV| \geq X \\ MA_{S_n} &= \frac{\sum_{i=1}^{NFS_n} ASMV}{NMB} \\ MA &= \frac{MA_{S_n}}{NVF} \end{aligned} \quad (4)$$

ここで、ASMVは大きさがX以上の動きベクトルのフレーム内絶対値総和、NMBはフレーム内での大きさがX以上の動きベクトルを持つマクロブロック数、NFS<sub>n</sub>はショット内の予測符号化フレーム数、NVFは1秒当たりのビデオフレーム数である。

【0031】動きアクティビティ判定部34において、求められた動きアクティビティMAと、ある指定された閾値THMAを比較し、MAがTHMAを超える場合に、該当するショット $S_n$ を概要情報SUMの候補から除外し、次のショット $S_{n+1}$ の入力処理へ移行する。一方、動きアクティビティ判定部34でショット $S_n$ のある単位時間における動きアクティビティMAが閾値THMAを超えない場合に、該当するショット $S_n$ を概要情報SUMの候補とする。

【0032】以上の処理により、オーディオビデオ情報AVを等間隔に分割した区間SAVにおいて、概要情報SUMの候補として選択されたショットSは概要情報登録部4に入力され、ショットSの情報が図6のショットメモリ41に保存される。次いで、区間SAVでの全ショットの処理が終了したか否かの判断が判断部42でなされ、この判断が否定の時には次のショットの処理が行われる。

【0033】一方、区間SAVの全ショットの処理が終了した時には、区間SAVにおける全てのショットSに関するショット長の総和SSLは、概要情報長判定部43において、各区間での平均概要情報長MSL (=SL/NP) に十分

12

近いかどうかを判断される。区間SAVでのショット長SSLと平均概要情報長MSLが十分に近いと見なされた場合には、区間SAVでの概要情報抽出処理を終了し、概要情報登録44で時間情報などが登録される。そして次の区間SAVn+1の入力処理へ移行する。このとき、代表ショット登録1Bで登録されていた代表ショットに関する各種特徴値はクリアされる。一方、SSLとMSLが近いと見なされない場合には、閾値変更部47で、概要情報SUMの候補としての判定に用いる一部または全ての閾値の値を変更し、それまでに抽出されている概要情報SUMの候補を対象として、SSLとMSLが十分近くなるまでショット長評価部16から動きアクティビティ評価部3での処理を再帰的に繰り返す。

【0034】これらの処理を、全ての区間SAVについて行い、最終的にオーディオビデオ情報AVの概要情報SUMを得て、概要情報SUMが登録される。このとき、概要情報の記述が指定されていれば、概要情報記述部5へ移行し、指定されていなければ処理を終了する。

【0035】概要情報記述部5では、図7に示されているように、概要情報SUMとして抽出された全てのショットSについて、概要情報記述部51でそれらの時間情報を少なくとも記述し、概要情報出力部52で概要情報記述ファイルとして出力する。記述するフォーマットとしては、例えばMPEG-7で規定されているフォーマットなどを用いることができる。全ての区間SAVについて記述が終了すると、全ての処理を終了する。また、概要情報SUMとして抽出された全てのショットSを結合して、別ファイルとして保存することができる。このとき、オーディオビデオファイルとして保存するか、オーディオとビデオを個別に保存することができる。

【0036】次に、前記した実施形態のオーディオビデオ概要情報抽出装置の機能は、ソフトウェア（プログラム）で実現することができる。該ソフトウェアは、光ディスク、フロッピー（登録商標）ディスク、ハードディスク等の記録媒体に記録することができる。

【0037】図8は、該記録媒体100に記録されるプログラムの一例を示すものであり、該記録媒体100には、圧縮オーディオビデオ情報のコンテンツ解析機能111、オーディオレベル評価機能112、動きアクティビティ評価機能113、概要情報登録機能114、および概要情報記述機能115が記録される。なお、該動きアクティビティ評価機能113は、省略してもよい。

【0038】また、前記コンテンツ解析機能111は、ビデオ情報をショットに分割する機能と、入力コンテンツをある基準に従って等間隔の区間に分割する機能と、該等間隔の区間に含まれるショットの長さを評価する機能および反復ショットを判定する機能の少なくとも一方とから構成することができる。また、前記オーディオレベル評価機能112は、無音を判定する機能、低レベル音を判定する機能、および高レベル音を判定する機能か

40

50

ら構成することができる。

【0039】また、前記動きアクティビティ評価機能113は、前記ショットに属するフレームの動きベクトルデータを抽出する機能と、該抽出された動きベクトルデータから動きアクティビティを計算する機能と、単位時間における動きアクティビティを計算する機能と、該単位時間における動きアクティビティを用いて概要情報の候補を判定する機能とから構成することができる。

【0040】また、抽出された概要情報を、オーディオビデオとして結合するか、またはオーディオとビデオ個別に結合するかをし、該結合した概要情報を別ファイルとして、記録媒体100に記録することができる。

【0041】なお、前記記録媒100には、ネットワークのように、データを一時的に記録保持するような伝送媒体も含まれる。

【0042】図9は、本発明の第2の実施形態であるオーディオ概要情報の抽出装置の構成を示すブロック図である。

【0043】まず、圧縮されたオーディオ情報CAが入力されると、サブバンドデータ抽出部A1でサブバンドデータSDを抽出する。抽出されたサブバンドデータSDは高レベル音評価部A3に送られる。サブバンドデータ抽出部A1の動作としては、第1の実施形態に示した無音・低レベル音評価部5におけるサブバンドデータ抽出部51と同様である。一方、非圧縮のオーディオ情報UAが入力されると、サブバンド解析部A2で入力オーディオがサブバンド解析され、解析された結果としてのサブバンドデータSDは同様に高レベル音評価部A3に送られる。

【0044】高レベル音評価部A3では、第1の実施形態に示した高レベル音解析部26に含まれるサブバンドエネルギー総和計算部262と、単位時間サブバンドエネルギー総和計算部263と同様の機能を持つ、図10のサブバンドエネルギー総和計算部A31と、単位時間サブバンドエネルギー総和計算部A32により、入力されたサブバンドデータSDから、それぞれサブバンドエネルギー総和SBEと単位時間でのサブバンドエネルギー総和SBが計算される。

【0045】次に、概要情報開始時間決定部A33において、単位時間サブバンドエネルギー総和SBが最大となる時間位置を、概要情報開始時間T\_startとして決定する。また、概要情報終了時間決定部A34では、単位時間サブバンドエネルギー総和SBが最大値の $\alpha$ 倍( $0 < \alpha < 1$ )となる時間位置を、概要情報終了時間T\_endとして決定する。このとき、 $T_{start} > T_{end}$ である。

【0046】概要情報登録部A4では、高レベル音評価部A3で決定された概要情報開始時間T\_startと、概要情報終了時間T\_endに基づいて概要情報を登録する。そして、概要情報記述部A5において上記時間情報を少なくとも記述し、概要情報記述ファイルとして出力する。記述するフォーマットとしては、例えばMPEG-7で規定さ

れているフォーマットなどを用いることができる。

【0047】オーディオ情報が複数存在する場合には、上記の処理を全てのオーディオ情報に対して行う。

【0048】前記した実施形態のオーディオ概要情報抽出装置の機能は、ソフトウェア(プログラム)で実現することができ、該ソフトウェアは、光ディスク、フロッピーディスク、ハードディスク等の記録媒体100に記録することができる。また、抽出されたオーディオ概要情報は、個別のファイルとして、該記録媒体100に記録することができる。

【0049】図11は、該記録媒体100に記録されるプログラムの一例を示すものであり、該記録媒体100には、サブバンドデータ抽出機能121または／およびサブバンド解析機能122、高レベル音評価機能123、概要情報登録機能124、および概要情報記述機能125が記録される。

【0050】図12は、本発明のオーディオビデオの概要情報再生装置の一実施形態を、構成図として表したものである。

【0051】前記手段により抽出されたオーディオビデオの概要情報SUMが入力されると、オーディオビデオ分離部P1において、概要情報のビデオ要素VSLUMとオーディオ要素ASUMに分離される。次に、ビデオ速度変換部P2では、外部から与えられた変換速度パラメータSPに従ってビデオ要素VSLUMを空間的に間引いてビデオの再生速度を変換する。同様にして、オーディオ要素ASUMはオーディオ速度変換部P3において変換速度パラメータSPに従ってビデオ要素VSLUMと同じ割合で時間的に間引かれ、オーディオの再生速度を変換する。オーディオの再生速度変換としては、例えばオーディオのフレームの周期的なスキップや、繰り返し再生と周期的スキップの組み合わせなどによって実現することができる。ここで、オーディオを1.5倍の速度にする場合、前者では

<再生するフレーム番号> 1, 2, 4, 5, 7, 8, 10, 11, …と連続する2フレームを再生し、次に続く1フレームスキップすることによって達成される。また後者では、

<再生するフレーム番号> 1, 1, 4, 4, 7, 7, 10, 10, …と同一フレームを2回繰り返して再生し、次に続く2フレームをスキップすることによって達成される。

【0052】速度を変換されたビデオ要素VSLUM及びオーディオ要素ASUMは、オーディオビデオ多重化・同期部P4に入力され、多重化及び同期処理が行われ、速度変換されたオーディオビデオの概要情報SUM'が得られる。得られたオーディオビデオの概要情報SUM'は、表示再生される。

【0053】

【発明の効果】以上の説明から明らかなように、本発明によれば、非圧縮または圧縮されたオーディオ情報、または圧縮されたオーディオビデオ情報に関して、それら

の内容を高速かつ効率的に把握するための概要情報を抽出することが可能になる。また、この抽出によって、オーディオビデオ情報の高速な閲覧が可能となる。

【0054】また、抽出される概要情報の長さは任意に指定することができると同時に、オーディオビデオ情報から均一に概要情報を抽出するため、コンテンツ全体の把握を効率的に行うことが可能となる。

【0055】また、概要情報に含まれる時間情報などを記述することにより、該当するオーディオビデオ情報の概要情報としての特徴記述を行うことが可能となり、コンテンツ記述の標準化であるMPEG-7などへも適用することが可能である。

【0056】また、抽出された概要情報の表示速度を変換するなどの、高機能な概要情報の再生を提供することが可能となる。

#### 【図面の簡単な説明】

【図1】 本発明の一実施形態の全体構成を示すブロック図である。

【図2】 図1のコンテンツ解析部の詳細構成を示すブロック図である。

【図3】 図1のオーディオレベル評価部の詳細構成を示すブロック図である。

【図4】 図3の高レベル音解析部の詳細構成を示すブロック図である。

【図5】 図1の動きアクティビティ評価部の詳細構成\*

\*を示すブロック図である。

【図6】 図1の概要情報登録部の詳細構成を示すブロック図である。

【図7】 図1の概要情報記述部の詳細構成を示すブロック図である。

【図8】 記録媒体に記録されるプログラムの概要を示す図である。

【図9】 本発明の他の実施形態のオーディオ概要情報抽出装置の構成を示すブロック図である。

10 【図10】 図8の高レベル音評価部の詳細構成を示すブロック図である。

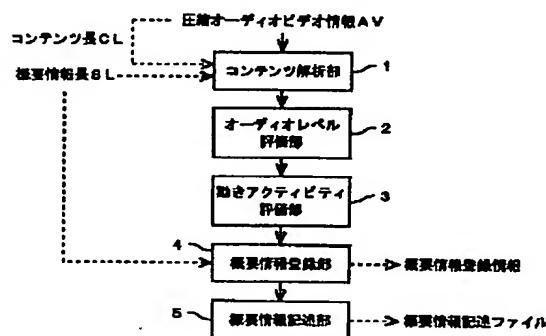
【図11】 記録媒体に記録されるプログラムの概要を示す図である。

【図12】 本発明の他の実施形態のオーディオ概要情報再生装置の構成を示すブロック図である。

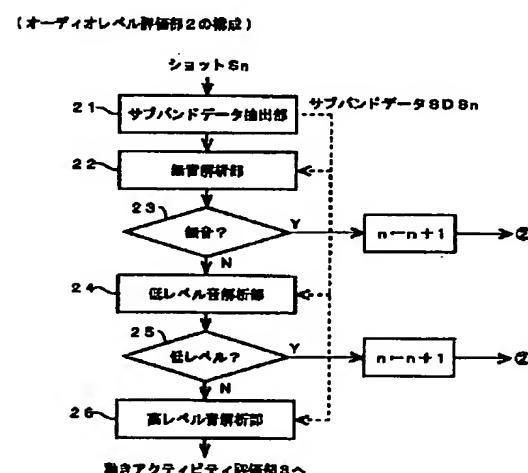
#### 【符号の説明】

1 … コンテンツ解析部、2 … オーディオレベル評価部、3 … 動きアクティビティ評価部、4 … 概要情報登録部、5 … 概要情報記述部、A1 … サブバンドデータ抽出部、A2 … サブバンド解析部、A3 … 高レベル音評価部、A4 … 概要情報登録部、A5 … 概要情報記述部、P1 … オーディオビデオ分離部、P2 … ビデオ速度変換部、P3 … オーディオ速度変換部、P4 … オーディオビデオ多重化・同期部、100 … 記録媒体。

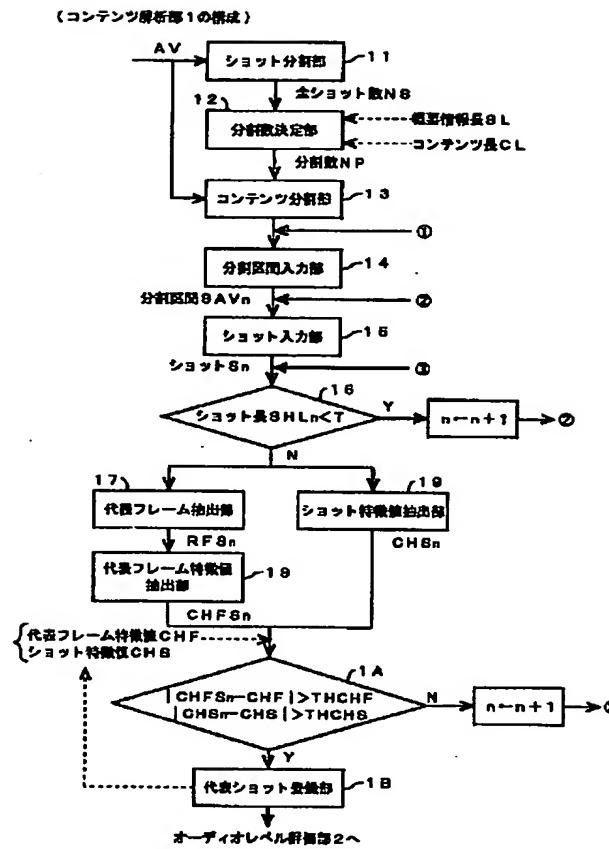
【図1】



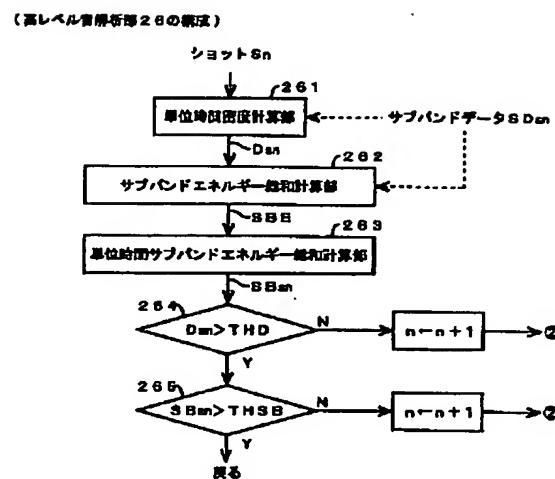
【図3】



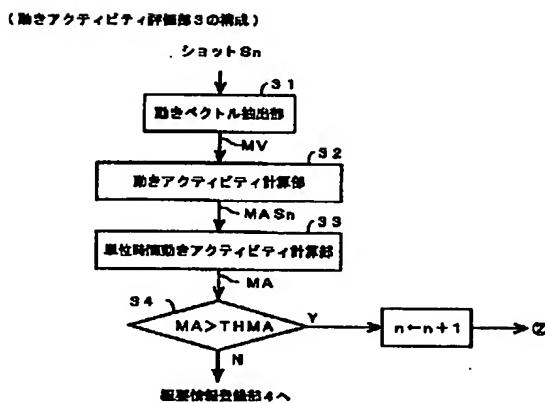
【図2】



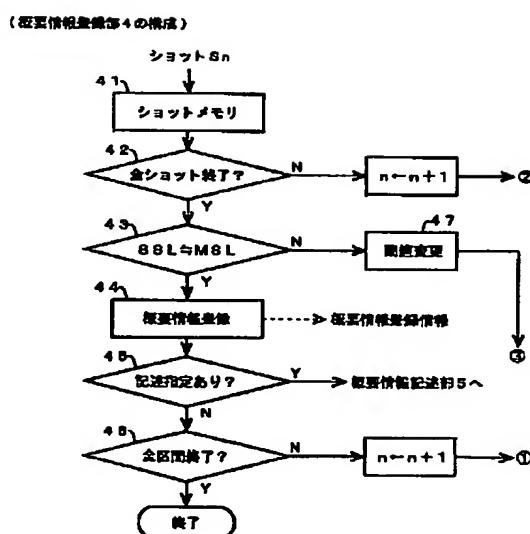
【図4】



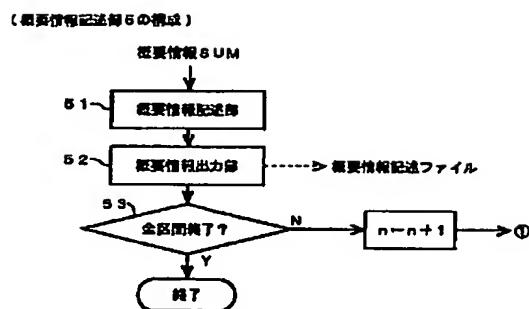
【図5】



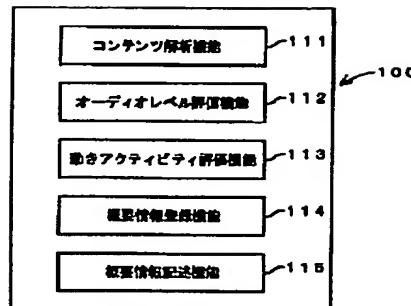
【図6】



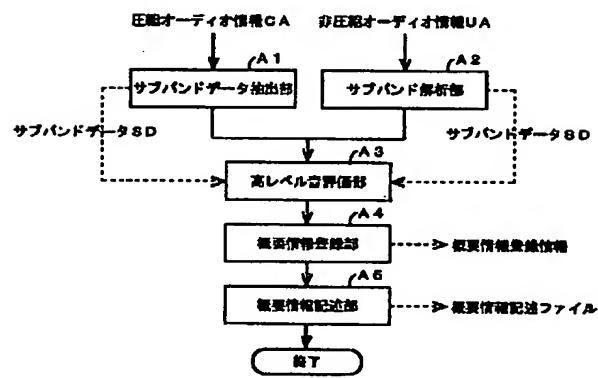
【図7】



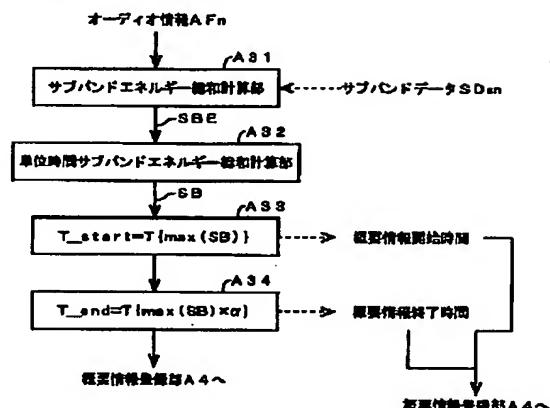
【図8】



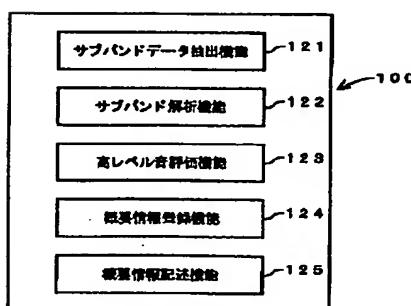
【図9】



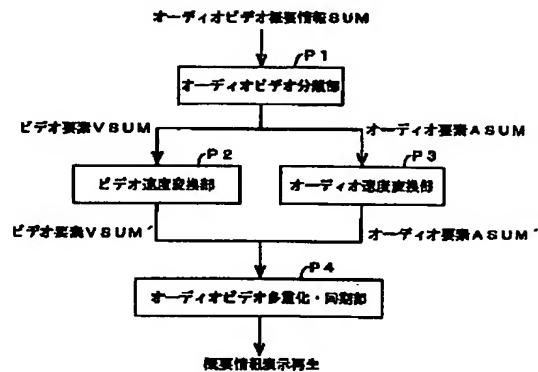
(高レベル音評価部A3の構成)



【図11】



【図12】



フロントページの続き

(72)発明者 柳原 広昌

埼玉県上福岡市大原2-1-15 株式会社  
ケイディディ研究所内

F ターム(参考) SC053 FA14 GB19 HA30 JA01  
SD044 AB02 BC01 BC03 CC04 DE91  
GK11 HL02  
SL096 BA20 GA08 GA59 HA04 HA08  
JA11